Statistical Bias in 3-D Reconstruction From a Monocular Video

Amit K. Roy-Chowdhury and Rama Chellappa, *Fellow, IEEE*

***Abstract*—The present state-of-the-art in computing the error statistics in three-dimensional (3-D) reconstruction from video concentrates on estimating the error covariance. A different source of error which has not received much attention is the fact that the reconstruction estimates are often significantly statistically biased. In this paper, we derive a precise expression for the bias in the depth estimate, based on the continuous (differentiable) version of structure from motion (SfM). Many SfM algorithms, or certain portions of them, can be posed in a linear least-squares (LS) framework $Ax = b$. Examples include initialization procedures for bundle adjustment or algorithms that alternately estimate depth and camera motion. It is a well-known fact that the LS estimate is biased if the system matrix A is noisy. In SfM, the matrix A contains point correspondences, which are always difficult to obtain precisely; thus, it is expected that the structure and motion estimates in such a formulation of the problem would be biased. Existing results on the minimum achievable variance of the SfM estimator are extended by deriving a *generalized* Cramer–Rao lower bound. A detailed analysis of the effect of various camera motion parameters on the bias is presented. We conclude by presenting the effect of bias compensation on reconstructing 3-D face models from rendered images.**

***Index Terms*—Correspondence errors, statistical bias, structure from motion (SfM).**

I. Introduction

STRUCTURE from motion (SfM) has been one of the most active research areas in computer vision for decades, with the result that today numerous algorithms exist which address various aspects of the problem (see [4], [8], and [14]). However, constructing accurate three-dimensional (3-D) models reliably from video sequences is still a challenging problem. Several researchers have analyzed the sensitivity and robustness of many of the existing algorithms, concentrating mostly on the error covariance of the depth estimates. A detailed review of existing work in the analysis of errors in SfM is available in one of our other papers [11]. A different source of error is the bias in depth estimation. Some authors, notably [1], have proved that there exists a bias in the translation and rotation estimates from stereo. In [7], Oliensis mentions correcting for the bias in the depth estimate, arising from the statistics of the parameters in their optimization function. However, quantification of the bias resulting from noisy image measurements is not discussed. Recently, it has been proposed that the bias in the optical flow field can be a possible explanation for many geometrical optical illusions [2]. A broad analysis of bias in 3-D reconstruction from different cues like motion, shape, and texture was presented in [5]. In this paper, we focus on quantifying the bias in the SfM problem and its implications for the design of practical 3-D modeling schemes. We show that the 3-D depth estimate obtained from SfM algorithms, using the differential optical flow based formulation, is statistically biased and, under many conditions, that the bias is numerically significant. We also show that the bias leads to a new lower bound on the minimum variance of an SfM estimate, thus extending the results in [13]. The effect of different camera motion parameters on the bias is studied. We present the effect of bias compensation on 3-D face reconstruction problems using images rendered from a texture-mapped model.

Manuscript received May 30, 2003; revised August 17, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hassan Foroosh.

A. K. Roy-Chowdhury is with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA.

R. Chellappa is with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742-3275 USA.

Digital Object Identifier 10.1109/TIP.2005.849775

II. Bias in Depth Reconstruction

A. Problem Formulation

Given two images I_1 and I_2, we are interested in computing the camera motion and structure of the scene from which these images were derived. If $p(x,y)$ and $q(x,y)$ are the horizontal and vertical velocity fields of a point (x,y) in the image plane, they are related to the 3-D object motion and scene depth (under the infinitesimal motion assumption) by

$$\begin{aligned} p(x,y) &= (-v_x + xv_z)g(x,y) + xy\omega_x - (1+x^2)\omega_y + y\omega_z \\ q(x,y) &= (-v_y + yv_z)g(x,y) + (1+y^2)\omega_x - xy\omega_y - x\omega_z \end{aligned} \tag{1}$$

where $\mathbf{V} = [v_x, v_y, v_z]$ and $\mathbf{\Omega} = [\omega_x, \omega_y, \omega_z]$ are the translational and rotational motion vectors, respectively, $g(x,y) = 1/Z(x,y)$ is the inverse scene depth, and all linear dimensions are normalized in terms of the focal length f of the camera. The problem is to estimate $\mathbf{V}, \mathbf{\Omega}$ and Z given (p,q). Equation (1) can be rewritten in a more useful form (because of the scale ambiguity [4]) as

$$\begin{aligned} p(x,y) &= (x - x_f)h(x,y) + xy\omega_x - (1+x^2)\omega_y + y\omega_z \\ q(x,y) &= (y - y_f)h(x,y) + (1+y^2)\omega_x - xy\omega_y - x\omega_z \end{aligned} \tag{2}$$

where $(x_f, y_f) = (v_x/v_z, v_y/v_z)$ is known as the *focus of expansion* (FOE) and $h(x,y) = v_z/Z(x,y)$.

For N points, the above equations can be represented using matrix notation, where the subscript is used to index the feature point. Consider N points (for a sparse depth map, this denotes

N feature points, while for a dense depth map it denotes the number of pixels in the image). Let us define

$$\begin{aligned}
\mathbf{h} &= (h_1, h_2, \ldots, h_N)^T_{N\times 1},\\
\mathbf{u} &= (p_1, q_1, p_2, q_2, \ldots, p_N, q_N)^T_{2N\times 1}\\
\mathbf{r}_i &= (x_i y_i, -(1+x_i^2), y_i)^T_{3\times 1},\\
\mathbf{s}_i &= (1+y_i^2, -x_i y_i, -x_i)^T_{3\times 1}\\
\mathbf{\Omega} &= (w_x, w_y, w_z)^T_{3\times 1},\\
\mathbf{Q} &= [\, r_1 \quad s_1 \quad r_2 \quad s_2 \quad \ldots \quad r_N \quad s_N \,]^T_{2N\times 3}\\
\mathbf{P} &= \left[\operatorname{diag}\begin{bmatrix} x_i - x_f \\ y_i - y_f \end{bmatrix}_{i=1,\ldots,N}\right]_{2N\times N}\\
\mathbf{B} &= [\mathbf{P} \quad \mathbf{Q}]_{2N\times(N+3)}\,.
\end{aligned} \tag{3}$$

Then, (2) can be written as

$$\mathbf{Bz} = \mathbf{u}, \text{ where } \mathbf{z}_{(N+3)\times 1} = \begin{bmatrix}\mathbf{h}^T & \mathbf{\Omega}^T\end{bmatrix}^T. \tag{4}$$

We want to compute $\mathbf{z}$ from $\mathbf{u}$. Consider the cost function which minimizes the reprojection error (i.e., bundle adjustment)

$$\begin{aligned}
C &= \sum_{i=1}^{N}[(p_i - \hat{p}_i)^2 + (q_i - \hat{q}_i)^2] = \frac{1}{2}\|\mathbf{Bz} - \mathbf{u}\|^2\\
&= \frac{1}{2}\sum_{i=1}^{N}(C_{pi}^2 + C_{qi}^2) = \frac{1}{2}\sum_{i=1}^{n} C_i^2
\end{aligned} \tag{5}$$

where $(\hat{p}_i, \hat{q}_i)$ are the projections of the depth and motion estimates, $\mathbf{z}$, onto the image plane and are obtained from the right hand side of the (2). In general, the above mentioned cost function requires nonlinear optimization, and various strategies (see [4] and [14]) have been proposed for this purpose.

B. Computation of Bias

Because of the fact that feature positions are never tracked perfectly, the 3-D reconstruction, in most situations, is statistically biased and the bias is significant. We give a precise expression for the bias and outline a proof in the Appendix .

As mentioned earlier, the solution of the cost function (5) involves nonlinear optimization which is extremely difficult, unless very good initial conditions are available ([4, Ch. 17]). The initial conditions are usually generated using a simpler method, e.g., a factorization approach. Also, different strategies are employed to solve this optimization problem. One of the common methods used is to first estimate the camera motion and then the depth. Another strategy is to update the camera motion and depth, one at a time, using the previous estimate of the other, until a convergence criterion is reached [taking advantage of the bilinear nature of the SfM (2)]. For monocular video sequences, it is often possible to first estimate the direction of motion (i.e., FOE) and then estimate the depth and rotational motion. The point to note is that if we assume an estimate of the camera motion or FOE and solve for the depth, we are essentially solving a linear system of equations. Similarly, if we assume that the depth is known and solve for the camera motion, we again have a linear system. This can be seen from the bilinear parameterization of (2). Also, the methods for generating initial conditions are often linear. It is a well known fact that the least-squares (LS) solution to a linear system of the form $Ax = b$ with errors in the system matrix A is biased [3]. When the SfM problem is posed in a LS framework, the matrix A involves the image coordinates, which almost always have measurement errors. Thus, it should be expected that the solution of the SfM problem would also have a bias. Such a bias is present if we adopt any of the above strategies to solve the nonlinear optimization problem using bundle adjustment or an initialization strategy that uses a linear LS criterion. To the best of our knowledge, this is the first attempt to explicitly compute and analyze the bias, arising from errors in feature tracking, in depth reconstruction from monocular video. Once the bias term is known, it can be compensated for and an unbiased estimate obtained at each of the above stages.

The actual value of the bias would be different for the different situations explained above. We consider the particular situation where the camera motion is known and derive the expression. We do this because one of the most common approaches to solving the 3-D reconstruction problem is to first estimate the camera motion, and then the depth. Expressions for the other conditions (e.g., simultaneous estimation of depth and rotation using a strategy that alternately estimates one of these parameters, or estimation of camera motion from the depth) can be similarly derived. For algorithms that estimate the parameters alternatively and update based on the previous estimates, the bias will propagate through the reconstruction strategy. For the case when the camera motion is known, (4) can be written as

$$\mathbf{b} = \mathbf{Ax} \tag{6}$$

with $\mathbf{A} \triangleq \mathbf{P}$, $\mathbf{x} \triangleq \mathbf{h}$ and $\mathbf{b} \triangleq \left[p_1 - \mathbf{r}_1^T\mathbf{\Omega}, q_1 - \mathbf{s}_1^T\mathbf{\Omega}, \ldots, p_N - \mathbf{r}_N^T\mathbf{\Omega}, q_N - \mathbf{s}_N^T\mathbf{\Omega}\right]^T$. We now state the main result of this paper in the form of a theorem.

Theorem 1: Consider the LS solution of (6), i.e., $\hat{x} = \hat{\mathbf{h}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$. For convenience, let us define

$$\begin{aligned}
\mathbf{M} \triangleq \mathbf{A}^T\mathbf{A} &= \operatorname{diag}\left[(x_i - x_f)^2 + (y_i - y_f)^2\right]\\
&\triangleq \operatorname{diag}[m_{ii}]_{i=1,\ldots,N} \qquad (7)\\
\mathbf{v} \triangleq \mathbf{A}^T\mathbf{b} &\triangleq [(x_i - x_f)v_{pi} + (y_i - y_f)v_{qi}]^T\\
&\triangleq [v_1, \ldots, v_N]^T \qquad (8)
\end{aligned}$$

where $v_{pi} = p_i - \mathbf{r}_i^T\mathbf{\Omega}$ and $v_{qi} = q_i - \mathbf{s}_i^T\mathbf{\Omega}$. The bias in the inverse depth estimate $\hat{\mathbf{h}}$ is $b(\hat{\mathbf{h}}) = E[\hat{\mathbf{h}}] - \bar{\mathbf{h}}$, where $\bar{\mathbf{h}}$ is the true value. If $\sigma_i^2 = E[\delta x_i^2] = E[\delta y_i^2]$ is the variance in the image coordinate measurements, then under the assumptions of the above formulation, the bias in the inverse depth estimate, $\hat{\mathbf{h}}_i$, of the i^{th} feature point is given by

$$\begin{aligned}
[\text{Bias}]_i &= b(\hat{\mathbf{h}}_i)\\
&= \frac{2\sigma_i^2}{m_{ii}^2}\left[(x_i - x_f)^2\mathbf{r}_{ix}^T + (y_i - y_f)^2\mathbf{s}_{iy}^T\right]\mathbf{\Omega}\\
&\quad + \frac{2\sigma_i^2}{m_{ii}^2}\left[(x_i - x_f)(y_i - y_f)(\mathbf{s}_{ix}^T + \mathbf{r}_{iy}^T)\right]\mathbf{\Omega}\\
&\quad + \frac{\sigma_i^2}{m_{ii}}\left[(x_i - x_f)\omega_y - (y_i - y_f)\omega_x - (\mathbf{r}_{ix}^T + \mathbf{s}_{iy}^T)\mathbf{\Omega}\right]
\end{aligned} \tag{9}$$

where f_{ix} represents the partial derivative of a function f_i with respect to x.

Proof: See Appendix A.

Further Comments : The bias can be similarly calculated for other parameters. For example, assume that the rotational motion parameters are unknown. In this case, $\mathbf{A} \triangleq \mathbf{Q}$, $\mathbf{x} \triangleq \mathbf{\Omega}$ and $\mathbf{b} \triangleq \mathbf{u} - \mathbf{Ph}$. If a strategy of alternately estimating depth and motion is adopted, the bias needs to be computed and compensated for at each stage (as explained later), else it will propagate through the different stages. The expression for the bias depends upon the variance σ_i^2. This can be estimated using the methods described in [11].

C. Analysis of Bias

The bias is a function of the camera motion parameters. It is most affected by the rotational motion of the camera. As can be seen from the expression in (9), the bias is negligibly small when the angular motion is zero (or very small). Once the structure and motion estimates are obtained, the bias can be computed and subtracted out of the estimate. For the biased estimate, $E[\hat{\mathbf{h}}] = \bar{\mathbf{h}} + b(\hat{\mathbf{h}})$, where $\bar{\mathbf{h}}$ is the true value. If $\hat{\mathbf{h}}_c = \hat{\mathbf{h}} - b(\hat{\mathbf{h}})$ is the bias compensated estimate, then $E[\hat{\mathbf{h}}_c] = E[\hat{\mathbf{h}}] - b(\hat{\mathbf{h}}) = \bar{\mathbf{h}}$, thus leading to an unbiased estimate. The effects of bias compensation and how it is dependent on the camera motion parameters in presented through simulations in Section IV.

The estimate of the two-frame bias in Theorem 1 can be extended to multiframe situations. Suppose that we have L two-frame reconstructions. Let $(\hat{\mathbf{h}}^1, \ldots, \hat{\mathbf{h}}^L)$ be the two-frame estimates aligned with respect to a particular frame of reference. Let the true value be $\bar{\mathbf{h}}$ and the bias in (9) be represented by $(b(\hat{\mathbf{h}}^1), \ldots, b(\hat{\mathbf{h}}^L))$, i.e., $E[\hat{\mathbf{h}}^i] = \bar{\mathbf{h}} + b(\hat{\mathbf{h}}^i)$, $i = 1, \ldots, L$. Assume that the estimates and the true value have the same scale (so that the problem of scale ambiguity does not arise). Then, the LS estimate for the structure over all L observations is $\hat{\mathbf{h}} = (1/L)\sum_{i=1}^{L} \hat{\mathbf{h}}^i$ (see [11] for details). Taking expectations on both sides, we see that the bias in the multiframe estimate is $b^L(\hat{\mathbf{h}}) = (1/L)\sum_{i=1}^{L} b(\hat{\mathbf{h}}^i)$, where $b(\hat{\mathbf{h}}^i)$ is obtained from (9) for the i and $(i+1)^{\text{st}}$ frame.

III. Bias-Modified Minimum Variance Bound for SfM

A lower bound on the variance of the SfM estimator, known as the Cramer–Rao lower bound (CRLB), was derived in [13] under Gaussian noise assumptions. This assumed the estimate to be unbiased. In the light of our discussion, this means that the *true* positions of the features are known, which is hardly the case ever. Since we know the expression for the bias, we can obtain a more accurate expression for the CRLB.

The general expression for the CRLB after incorporating the bias in the estimate and under the proper regularity assumptions is [12]

$$\Sigma_\theta(g) \geq b_\theta(g) b_\theta(g)^T + (I_{p\times p} + \nabla_\theta b_\theta(g)) M^{-1}(\theta)(I_{p\times p} + \nabla_\theta b_\theta(g))^T \quad (10)$$

where g is the estimate of the parameter θ, b is the bias of the estimate, and M the Fisher information matrix. $I_{p\times p}$ is an identity matrix and ∇_θ is the gradient with respect to θ.

Let $\hat{\mathbf{h}}$ denote the estimate of $\bar{\mathbf{h}}$ (the true value). Let the bias in the multiframe estimate be denoted by $b(\hat{\mathbf{h}})$ (Section (II)). It can be shown that the variance of the biased estimate $\hat{\mathbf{h}}$, represented as $\Sigma(\hat{\mathbf{h}}) = E\left[(\hat{\mathbf{h}} - \bar{\mathbf{h}})(\hat{\mathbf{h}} - \bar{\mathbf{h}})^T\right]$ can be expressed as $\Sigma(\hat{\mathbf{h}}) = E\left[(\hat{\mathbf{h}} - E[\hat{\mathbf{h}}])(\hat{\mathbf{h}} - E[\hat{\mathbf{h}}])^T\right] + b(\hat{\mathbf{h}}) b^T(\hat{\mathbf{h}})$. In [11], we showed how to compute the covariance represented by the first term on the right hand side of the above equation. This was done using the implicit function theorem and did not require assumptions on the distributions of the noise in the observations (i.e., feature positions). Adding the bias to this expression, we can now compute a general expression for the covariance of the inverse depth estimate. Since the bias does not depend on $\bar{\mathbf{h}}$, $\nabla_{\bar{\mathbf{h}}} b(\hat{\mathbf{h}}) = 0$. Let the FI matrix for the inverse depth parameter be denoted by $F(\bar{\mathbf{h}})$. By substituting the values of the different terms in (10), the variance of the inverse depth estimate is lower bounded as

$$\Sigma(\hat{\mathbf{h}}) \geq b(\hat{\mathbf{h}}) b^T(\hat{\mathbf{h}}) + F^{-1}(\bar{\mathbf{h}}). \quad (11)$$

This is the expression of the generalized CRLB of the inverse depth, obtained from the optical flow based SfM equations in (2). Under the special case of additive Gaussian noise assumptions in p and q in (2), the FI matrix is as derived in [13]. Substituting that expression into (11), we obtain the generalized CRLB for the Gaussian noise case.

IV. Experimental Results

The first set of experiments deals with a set of 50 3-D points whose true positions are known. The initial positions of these points were set randomly. Different kinds of motion were applied to these points so as to obtain various motion tracks for each of them. The perspective projections of these points were generated on a 512×512 pixel grid. This resulted in creating a set of tracked features. The median value of the true motion between two consecutive frames (median computed over all frames and features) was around 15 pixels in both the horizontal and vertical directions.

Effect of Bias on Reconstruction: A set of ten random feature points, chosen from the above set, were tracked across a few frames. The depths from each pair of frames were obtained and then fused together using stochastic approximation, as explained in [11]. To fix the scale of the reconstruction, the depth at the first point was used. In these experiments we considered the case of nonzero but constant translational and rotational camera motion. The effect of measurement noise was studied by adding noise to the feature positions. Fig. 1(a) shows the effect of bias compensation with noise of standard deviation $\sigma_x = 5$ pixels. It can be seen that bias compensation makes the estimate closer to the true value (i.e., reduces the bias) and gives significant advantages for some of the points.

Variation of Bias With Individual Camera Motion Parameters: In this set of experiments, we analyzed the effects of the camera motion through numerical simulations. The focal length of the camera was assumed to be known. Each of the six motion parameters was varied over a certain range of values, keeping all the others fixed. The range for the variation in the different experiments was as follows: a) $v_x \in (0, 1)$ cm/frame; b) $v_y \in$

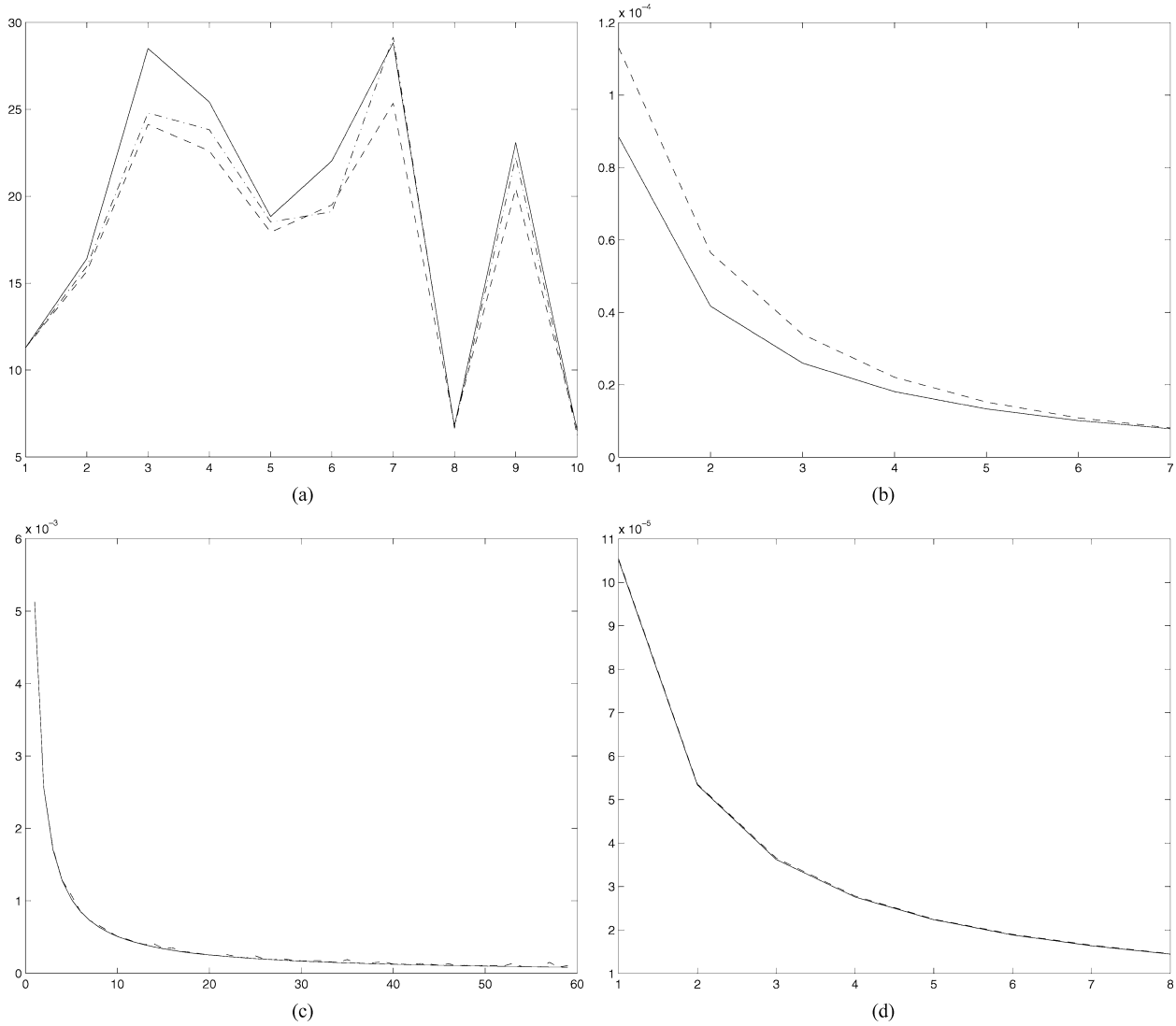


Fig. 1. (a) Reconstruction plot for noise with standard deviation $\sigma_x = 5$ pixels. The plot is for a set of ten 3-D points tracked over 15 frames. The camera is moving with constant, nonzero translation and rotation. The solid lines indicate the true depth values, the dashed lines indicate reconstruction without bias compensation, and the dashed and dotted lines indicate reconstruction with bias compensation. (b)-(d) CRLB of the inverse depth as a function of the number of frames, for different camera motion parameters, which are as follows: (b) $x_f = 10$, $y_f = 10$, $\omega_x = \omega_y = \omega_z = 1$ degree/frame; (c) $x_f = 1$, $y_f = 1$, $\omega_x = \omega_y = \omega_z = 0$; (d) uniform acceleration, $\omega_x = \omega_y = \omega_z = 0$. (a) Depth versus number of points. (b) CRLB versus number of frames (c) CRLB versus number of frames. (d) CRLB versus number of frames.

$(0, 1)$ cm/frame; c) $v_z \in (0, 1)$ cm/frame; d) $\omega_x \in (0, 10)$ degrees/frame; e) $\omega_y \in (0, 10)$ degrees/frame; f) $\omega_z \in (0, 10)$ degrees/frame. The motion terms which affect the bias the most (approximately $\pm$ 5%–10% of the true depth) are (v_x, v_y, ω_x, and ω_y). The effect of (v_z, ω_z) on bias is almost negligible.

Bias-Modified Generalized CRLB: The generalized CRLB, derived in Section III, varies with the nature of the camera motion. From Fig. 1(b), we see that there is a distinct upward shift in the minimum reconstruction error for nonzero translation and rotation. When the rotational velocity is zero, the bias term is negligible; the difference in the CRLB is too small to represent in the plots of Fig. 1(c)–(d). This is the case irrespective of the actual value of $\mathbf{V}$, including any acceleration. The reason for this can be understood from the expression for the bias in (9). The only term in (9) that contributes to the bias if $\mathbf{\Omega} = 0$ is $(\sigma_i^2/m_{ii})\left[(x_i - x_f)\omega_y - (y_i - y_f)\omega_x\right]$. This would be small unless the variances of the errors in the feature positions are very large, in which case, the solution itself would be unreliable.

The conclusion that can be drawn from this analysis is that the parameters that affect bias the most are the camera angular motion values. *For zero (or close to zero) rotation, the bias in the estimate is negligible.* While this can be understood mathematically from the expression for the bias as derived above, a physical explanation is a topic for future study.

A. Bias in 3-D Face Modeling

For this particular problem, we decided to use the database available on the World Wide Web at http://sampl.eng.ohio-state.edu/~sampl/data/3DDB/RID/minolta/faces-hands.1299/index.html. This database includes the 3-D depth model obtained from a range scanner and a frontal image of the face for texture mapping. We used the 3-D model and the texture map to create a sequence of images after specifying the camera

TABLE I
EFFECT OF BIAS ON 3-D FACE RECONSTRUCTION

Subject Index	Peak % Bias	Avg. % Error (before bias compensation)	Avg. % Error (after bias compensation)
1 (frame 001)	30	3.8	3.6
2 (frame 002)	34	3.2	3.1
3 (frame 003)	29	3.0	2.7
4 (frame 004)	21	3.9	3.7
5 (frame 005)	26	3.2	3.0

motion. The camera motion consisted of translation along the x and z axes and rotation about the y axis. Given this sequence of images, we estimate the 3-D model using a 3-D face reconstruction algorithm [10] (not explained in this paper) and the bias in the reconstruction using the (9). We present here the results on the first five face models in the above mentioned database. Following the convention on the website, we refer to the five subjects as "frame001" to "frame005." From Table I, we see that the peak value of the bias is a significant percentage of the true depth value. This happens only for a few points; however, it has significant impact on the 3-D face model because of interpolation techniques which, invariably, are a part of any method to build 3-D models. The third and fourth columns in Table I represent the root mean square (RMS) error of the reconstruction represented as a percentage of the true depth and calculated before and after bias compensation. The change in the average error after bias compensation is very small. However, by itself, this number is misleading. The average error in reconstruction may be small, however, even one outlier has the potential to create a very poor reconstruction. Hence, it is very important to compensate for the bias in problems related to 3-D reconstruction from a monocular video. It is justified to ask whether these few points at which the bias is large could have been removed by an outlier rejection method on the output 3-D model. Even if that is possible, the bias estimation and compensation technique described in this paper can prevent the cause of these erroneous points, as well as provide a physical understanding for their reason of occurrence.

V. CONCLUSION

Traditionally, the analysis of the accuracy of 3-D reconstruction has focused on the error covariance of the estimate. In this paper, we have pointed out that there is another source of error in the SfM problem, namely the bias in the estimate. Our derivation of the bias term was based on the fact that the solution of a LS estimation problem with noisy system matrix is statistically biased. The SfM problem or certain stages of existing algorithms can be posed in a linear LS framework. The system matrix in these formulations contains the positions of the features, that can never be obtained exactly. A new minimum error bound (i.e., generalized CRLB) for SfM was proposed after incorporating the bias term. Simulations were carried out in order to show the effects of the different camera motion parameters on the bias. It was observed that the bias is negligibly small if the camera angular motion is small.

APPENDIX
OUTLINE OF PROOF OF THEOREM 1

We give a brief outline of the proof. The details can be found in [9].

Expanding $\hat{\mathbf{h}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$ in a Taylor series around the true value $\bar{\mathbf{h}}$, (i.e., the noise $N = 0$) and assuming the mean deviation in that region to be zero (i.e., $E[\delta x_i] = E[\delta y_i] = E[\delta x_f] = E[\delta y_f] = 0$), and all the components δx_i, δy_i, δx_f, δy_f to be mutually uncorrelated, we can express

$$E[\hat{\mathbf{h}}] \approx \bar{\mathbf{h}} + \sum_{i=1}^{N}\left[\frac{\partial^2\hat{\mathbf{h}}}{\partial\delta x_i^2}E\left[\frac{\delta x_i^2}{2}\right] + \frac{\partial^2\hat{\mathbf{h}}}{\partial\delta y_i^2}E\left[\frac{\delta y_i^2}{2}\right]\right] + \frac{\partial^2\hat{\mathbf{h}}}{\partial\delta x_f^2}E\left[\frac{\delta x_f^2}{2}\right] + \frac{\partial^2\hat{\mathbf{h}}}{\partial\delta y_f^2}E\left[\frac{\delta y_f^2}{2}\right] \quad (12)$$

where all the partials are computed at $N = 0$. The above equation is actually a simplified version which does not take into account the errors in $(p_i, q_i, \omega_X, \omega_Y, \omega_Z)$. This is because

$$\frac{\partial^2\hat{\mathbf{h}}}{\partial\delta p_i^2} = \frac{\partial^2\hat{\mathbf{h}}}{\partial\delta q_i^2} = \frac{\partial^2\hat{\mathbf{h}}}{\partial\delta\omega_x^2} = \frac{\partial^2\hat{\mathbf{h}}}{\partial\delta\omega_y^2} = \frac{\partial^2\hat{\mathbf{h}}}{\partial\delta\omega_z^2} = 0.$$

In the absence of any measurement noise, the expected value of the estimate obtained from the LS solution should equal the true value $\bar{\mathbf{h}}$. However, since there exist errors in the measurement model, the estimate is biased and the sum of the last four terms on the right hand side of (12) represents the total bias in the estimate. In order to calculate the bias, we need to compute the derivatives in (12). We can compute all the derivatives using the fact that for an arbitrary matrix Q, $-\partial Q^{-1}/\partial x = Q^{-1}(\partial Q/\partial x)Q^{-1}$ [6]. The final result can be obtained by substituting substituting the partials in (12).

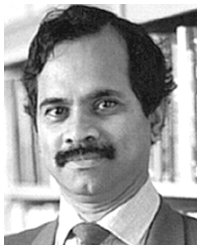
REFERENCES

[1] Y. Aloimonos, Ed., *Visual Navigation*. Hillsdale, NJ: Lawrence Erlbaum, 1996, pp. 61–88.
[2] C. Fermuller, D. Shulman, and Y. Aloimonos, "The statistics of optical flow," *Comput. Vis. Image Understand.*, vol. 82, pp. 1–32, 2001.
[3] W. A. Fuller, *Measurement Error Models*. New York: Wiley, 1987.

[4] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[5] H. Ji and C. Fermuller, "Uncertainty in 3D shape estimation," presented at the ICCV Workshop on Statistical and Computational Theories of Vision, 2003.
[6] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.
[7] J. Oliensis, "A multi-frame structure-from-motion algorithm under perspective projection," *Int. J. Comput. Vis.*, vol. 34, pp. 1–30, Aug. 1999.
[8] ——, "A critique of structure-from-motion algorithms," *Comput. Vis. Image Understand.*, vol. 80, no. 2, pp. 172–214, Nov. 2000.
[9] A. Roy Chowdhury, "Statistical analysis of 3D modeling from monocular video streams," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Maryland, College Park, 2002.
[10] A. Roy Chowdhury and R. Chellappa, "Face reconstruction from monocular video using uncertainty analysis and a generic model," *Comput. Vis. Image Understand.*, pp. 188–213, Jul.-Aug. 2003.
[11] ——, "Stochastic approximation and rate-distortion analysis for robust structure and motion estimation," *Int. J. Comput. Vis.*, pp. 27–53, Oct. 2003.
[12] J. Shao, *Mathematical Statistics*. New York: Springer-Verlag, 1998.
[13] G. S. Young and R. Chellappa, "Statistical analysis of inherent ambiguities in recovering 3-d motion from a noisy flow field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 10, pp. 995–1013, Oct. 1992.
[14] Z. Zhang and O. Faugeras, *3D Dynamic Scene Analysis*. New York: Springer-Verlag, 1992.

Amit K. Roy-Chowdhury received the B.S. degree in electrical Eengineering from Jadavpur University, Calcutta, India, in 1985, the M.S. degree in engineering in systems science and automation from the Indian Institute of Science, Bangalore, in 1997, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Maryland, College Park, in 2002, where he worked on statistical error characterization of 3-D modeling from monocular video sequences.

He is an Assistant Professor in the Electrical Engineering Department, University of California, Riverside. He was previously with the Center for Automation Research, University of Maryland, as a Research Associate. He was involved in projects related to face, gait, and activity modeling and recognition. His research interests are in signal, image and video processing, computer vision, and pattern recognition.



Rama Chellappa (S'78–M'79–SM'83–F'92) received the B.E. (Hons.) degree from the University of Madras, Madras, India, in 1975 and the M.E. (Distinction) degree from the Indian Institute of Science, Bangalore, in 1977. He received the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1978 and 1981, respectively.

Since 1991, he has been a Professor of electrical engineering and an affiliate Professor of Computer Science at the University of Maryland, College Park. He is also affiliated with the Center for Automation Research (Director) and the Institute for Advanced Computer Studies (permanent member). Prior to joining the University of Maryland, he was an Assistant Professor (1981 to 1986) and an Associate Professor (1986 to 1991) and Director of the Signal and Image Processing Institute (1988 to 1990) with the University of Southern California (USC), Los Angeles. Over the last 22 years, he has published numerous book chapters and peer-reviewed journal and conference papers. He has edited a collection of Papers on Digital Image Processing (Los Alamitos, CA: IEEE Computer Society Press, 1992), coauthored a research monograph on *Artificial Neural Networks for Computer Vision* (with Y. T. Zhou) (New York: Springer-Verlag, 1990), and co-edited a book on *Markov Random Fields: Theory and Applications* (with A. K. Jain) (New York: Academic, 1993). His current research interests are face and gait analysis, 3-D modeling from video, automatic target recognition from stationary and moving platforms, surveillance and monitoring, hyperspectral processing, image understanding, and commercial applications of image processing and understanding.

Dr. Chellappa has served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE TRANSACTIONS ON NEURAL NETWORKS. He was Co-Editor-in-Chief of *Graphical models and Image Processing*. He is now serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. He served as a member of the IEEE Signal Processing Society Board of Governors from 1996 to 1999. Currently, he is serving as the Vice President of Awards and Membership for the IEEE Signal Processing Society. He has served as a General the Technical Program Chair for several IEEE international and national conferences and workshops. He received several awards, including the National Science Foundation (NSF) Presidential Young Investigator Award, an IBM Faculty Development Award, the 1990 Excellence in Teaching Award from School of Engineering at USC, the 1992 Best Industry Related Paper Award from the International Association of Pattern Recognition (with Q. Zheng), and the 2000 Technical Achievement Award from the IEEE Signal Processing Society. He was elected as a Distinguished Faculty Research Fellow (1996 to 1998) at the University of Maryland, he is a Fellow of the International Association for Pattern Recognition, and he received a Distinguished Scholar-Teacher award from the University of Maryland in 2003.